

Clase 16: Intervalos de confianza para diferencia de medias con muestras normales y no-normales

Universidad Nacional de Colombia - Sede Medellín

Caso 1: IC para $\mu_X - \mu_Y$ con poblaciones normales y varianzas conocidas

Suponga que se tiene una m.a. X_1, \dots, X_n de una distribución $N(\mu_X, \sigma_X^2)$ y que Y_1, \dots, Y_m es otra m.a. de una distribución $N(\mu_Y, \sigma_Y^2)$ y que ambas m.a son independientes entre sí.

Como ambas muestras son E.I. entonces \bar{X} y \bar{Y} también lo son. Ahora

$$E[\bar{X} - \bar{Y}] = \mu_X - \mu_Y \quad \text{y} \quad \text{Var}[\bar{X} - \bar{Y}] = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

Debido a que \bar{X} y \bar{Y} son normales, entonces $\bar{X} - \bar{Y}$ también lo es:

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0, 1).$$

Si σ_X^2 y σ_Y^2 **son conocidas**, el IC para $\mu_X - \mu_Y$ está dado por:

$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}.$$

Caso 2: IC para $\mu_X - \mu_Y$ con poblaciones normales y varianzas desconocidas pero iguales $\sigma_X^2 = \sigma_Y^2$

Para hallar un IC al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ se debe usar el hecho de que:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2},$$

donde

$$S_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}.$$

Así, un IC de confianza al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es de la forma:

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}.$$

Caso 3: IC para $\mu_X - \mu_Y$ con poblaciones normales y varianzas desconocidas pero diferentes $\sigma_X^2 \neq \sigma_Y^2$

En este caso se tiene que

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} \sim t(v),$$

donde

$$v = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m-1}}.$$

Así, un IC al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es:

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, v} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}.$$

IC para $\mu_X - \mu_Y$ con poblaciones no normales con tamaños de muestra grande

Suponga que X_1, \dots, X_n es una m.a. de una población con media μ_X y varianza σ_X^2 . Sea Y_1, \dots, Y_m otra m.a. independiente de la anterior de otra población con media μ_Y y varianza σ_Y^2 .

¿Como calcular un IC al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$?

Como ambas muestras son E.I. entonces \bar{X} y \bar{Y} también lo son. Ahora:

$$E[\bar{X} - \bar{Y}] = \mu_X - \mu_Y \quad \text{y} \quad \text{Var}[\bar{X} - \bar{Y}] = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

Suponiendo tamaños de muestras mayores a 30, por el T.L.C.,

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \stackrel{n, \widetilde{m \rightarrow +\infty}}{\text{aprox}} N(0, 1).$$

Caso 4: IC para $\mu_X - \mu_Y$ con poblaciones no normales, varianzas conocidas, n y m grandes

Si las varianzas son conocidas, un IC aproximado al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es de la forma:

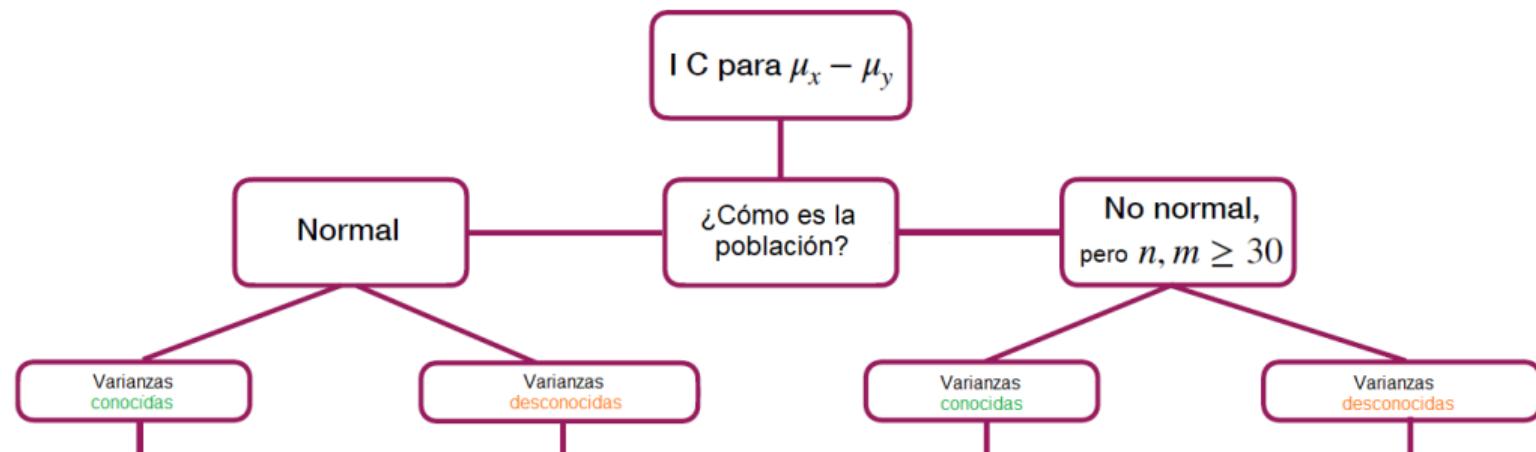
$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}.$$

Caso 5: IC para $\mu_X - \mu_Y$ con poblaciones no normales, varianzas desconocidas, n y m grandes

Si σ_X^2 y σ_Y^2 son desconocidas, podemos usar s_X^2 y s_Y^2 las respectivas varianzas muestrales. Así, un IC aproximado al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es

$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}.$$

Resumen IC para $\mu_x - \mu_y$



$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

Asumiendo varianzas iguales:

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

$$\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$$

Asumiendo varianzas diferentes

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, v} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$$

IC unilateral superior para $\mu_X - \mu_Y$

Similarmente al caso del IC para la media, se tiene que:

Un Intervalo de Confianza Unilateral superior al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es de la forma:

$$\left(\bar{x} - \bar{y} - z_{\alpha} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, +\infty \right)$$

$$\left(\bar{x} - \bar{y} - z_{\alpha} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}, +\infty \right)$$

$$\left(\bar{x} - \bar{y} - t_{\alpha,v} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}, +\infty \right)$$

$$\left(\bar{x} - \bar{y} - t_{\alpha,n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}, +\infty \right)$$

IC unilateral inferior para $\mu_X - \mu_Y$

Un Intervalo de Confianza Unilateral inferior al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es de la forma:

$$\left(-\infty, \bar{x} - \bar{y} + z_\alpha \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \right)$$

$$\left(-\infty, \bar{x} - \bar{y} + z_\alpha \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} \right)$$

$$\left(-\infty, \bar{x} - \bar{y} + t_{\alpha,v} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} \right)$$

$$\left(-\infty, \bar{x} - \bar{y} + t_{\alpha,n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

Ejemplo de IC para $\mu_X - \mu_Y$

En un proceso químico pueden emplearse dos catalizadores 1 y 2. Se sospecha que el catalizador 2 podría tener en promedio mejor rendimiento que el catalizador 1. Para verificar esta afirmación, se prepararon 36 lotes con el catalizador 1 y 49 con el catalizador 2. Ambas muestras son aleatorias. Para el primer catalizador el rendimiento promedio fue del 86% con una desviación estándar del 3% y para el segundo catalizador se obtuvo un rendimiento promedio de 89% con una desviación estándar de 2%. ¿Son ciertas las sospechas? Considere $\alpha = 0.05$.

Solución

Sea X_1, \dots, X_{36} que representa los rendimientos con el catalizador 1.

Sea Y_1, \dots, Y_{49} que representa los rendimientos con el catalizador 2.

Ambas m.a. son E.I.

Como las dos m.as no provienen de poblaciones normales y σ_X^2, σ_Y^2 son desconocidas, entonces, un IC aproximado unilateral inferior al $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es:

$$\left(-\infty, \bar{x} - \bar{y} + z_\alpha \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} \right).$$

De la información muestral se tiene que:

$$\bar{x} = 86, \quad s_x = 3, \quad n = 36; \quad \bar{y} = 89, \quad s_y = 2, \quad m = 49.$$

Si $\alpha = 0.05$, entonces $z_\alpha = z_{0.05} = 1.645$. Con esto se tiene que un IC aproximado al 95% para $\mu_X - \mu_Y$ es:

$$\left(-\infty, (86 - 89) + 1.645 \sqrt{\frac{3^2}{36} + \frac{2^2}{49}} \right) \Leftrightarrow$$
$$\left(-\infty, -3 + 1.645 \sqrt{\frac{65}{196}} \right) \Leftrightarrow (-\infty, -2.0527).$$

Se concluye que $\mu_X - \mu_Y < 0$, con una confianza aproximada del 95%, es decir, $\mu_Y > \mu_X$; con esto, las sospechas son bien soportadas por la información muestral: el rendimiento promedio del catalizador 2 es superior al del catalizador 1.

Ejemplo de IC para $\mu_X - \mu_Y$

Se hicieron pruebas de resistencia a la tensión (kg/mm^2) a dos tipos distintos de varillas para alambres y se obtuvieron los siguientes resultados:

<u>Grado</u>	<u>Tamaño Muestral</u>	<u>Media Muestral</u>	<u>Desviación Estándar</u>
AISI - 1064	n = 129	107.6	1.3
AISI - 1078	m = 129	123.6	2.0

¿Indican estos datos que la resistencia promedio real del grado 1078 supera a la del grado 1064 en más de $10 kg/mm^2$? Comente.

Solución

Sea X_1, \dots, X_{129} una m.a. que representa las resistencias de las 129 varillas tipo 1064.

Suponga que

$E[X_i] = \mu_X$ y $Var[X_i] = \sigma_X^2$; $i = 1, 2, \dots, 129$. Sea Y_1, \dots, Y_{129} otra m.a. que representa las resistencias de las 129 varillas tipo 1078. Asuma que $E[Y_j] = \mu_Y$ y $V[Y_j] = \sigma_Y^2$; $j = 1, 2, \dots, 129$. Ambas m.a. E.I.

Se quiere ver si $\mu_Y - \mu_X > 10$, para ello hallaremos un IC aproximado unilateral superior al 95% para $\mu_Y - \mu_X$.

Un IC aproximado unilateral superior al 95% para $\mu_Y - \mu_X$ es:

$$\left(\bar{y} - \bar{x} - z_{\alpha} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}, +\infty \right).$$

La información muestral indica que

$$\bar{x} = 107.6, \quad s_X = 1.3, \quad \bar{y} = 123.6, \quad s_Y = 2.0, \quad z_{0.05} = 1.645.$$

Se tiene que un IC aproximado unilateral superior al 95% para $\mu_Y - \mu_X$ es:

$$\left((123.6 - 107.6) - 1.645 \sqrt{\frac{1.3^2}{129} + \frac{2.0^2}{129}}, +\infty \right) \Leftrightarrow (15.6545, +\infty).$$

Observe que con una confianza aproximada del 95%, se tiene que $\mu_Y - \mu_X > 10$, es decir, la sospecha es cierta.

Ejemplo de IC para $\mu_X - \mu_Y$

Se investiga el diámetro de las varillas de acero fabricadas en dos diferentes máquinas de extrusión. Para ello se toman dos m.as de tamaños $n = 14$ y $m = 18$ de poblaciones normales. Las medias y varianzas muestrales son:

- Media muestral y varianza muestral de los diámetros de las varillas provenientes de la máquina 1: 8.73 y 0.35, respectivamente.
- Media muestral y varianza muestral de los diámetros de las varillas provenientes de la máquina 2: 8.68 y 0.4, respectivamente.

Construya un IC bilateral al 95% para la diferencia en los diámetros promedio de las varillas fabricadas por las dos máquinas, asumiendo:

- a) $\sigma_X^2 = \sigma_Y^2$
- b) $\sigma_X^2 \neq \sigma_Y^2$

Solución

Suponga que X_1, \dots, X_{15} es una m.a que representa los diámetros de las varillas producidas por la máquina 1; sea Y_1, \dots, Y_{18} otra m.a que representa los diámetros de las varillas producidas por la máquina 2. Además:

$$X_i \sim N(\mu_X, \sigma_X^2), \quad Y_j \sim N(\mu_Y, \sigma_Y^2).$$

De la información muestral se tiene que:

$$\bar{x} = 8.73, \quad s_X^2 = 0.35, \quad n = 14, \quad \bar{y} = 8.68, \quad s_Y^2 = 0.4 \quad y \quad m = 18.$$

- a) Ambas poblaciones normales, muestras pequeñas y σ_X^2, σ_Y^2 , desconocidas pero iguales. Un IC bilateral al 95% para $\mu_X - \mu_Y$ es:

$$(8.73 - 8.68) \pm t_{0.025,30} S_p \sqrt{\frac{1}{14} + \frac{1}{18}},$$

$$S_p^2 = \frac{13(0.35) + 17(0.4)}{30} = 0.3783, \quad S_p = 0.6151, \quad t_{0.025,30} = 2.042.$$

Un IC al 95% para $\mu_X - \mu_Y$ es:

$$(8.73 - 8.68) \pm 2.042 \times 0.6151 \sqrt{\frac{1}{14} + \frac{1}{18}} \Leftrightarrow (-0.3976, 0.497).$$

Este intervalo incluye el cero y permite concluir que $\mu_X - \mu_Y = 0$, es decir, no existe diferencias significativas en los diámetros medios de las varillas producidas por ambas máquinas.

- b) Ambas poblaciones normales, muestras pequeñas y σ_X^2, σ_Y^2 , desconocidas pero diferentes. Un IC bilateral al 95% para $\mu_X - \mu_Y$ es:

$$(8.73 - 8.68) \pm t_{0.025,v} \sqrt{\frac{s_X^2}{14} + \frac{s_Y^2}{18}}$$

$$\text{Ahora, como } v = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m-1}} = \frac{\left(\frac{0.35}{14} + \frac{0.4}{18}\right)^2}{\frac{\left(\frac{0.35}{14}\right)^2}{13} + \frac{\left(\frac{0.4}{18}\right)^2}{17}} = 28.$$

$$t_{0.025,28} = 2.05$$

Se tiene que un IC al 95% para $\mu_X - \mu_Y$ es:

$$(8.73 - 8.68) \pm 2.05 \sqrt{\frac{0.35}{14} + \frac{0.4}{18}} \Leftrightarrow (-0.3954, 0.4954).$$

Este intervalo incluye el cero y permite concluir que $\mu_X - \mu_Y = 0$. No existe diferencias significativas en los diámetros de las varillas fabricadas en ambas máquinas de extrusión.

Ejemplo de IC para $\mu_X - \mu_Y$

Se realizó un experimento para comparar el tiempo promedio requerido por el cuerpo humano para absorber dos medicamentos, A y B. Se cree que el fármaco B se absorbe en promedio más rápido que el A. Para verificarlo se eligieron al azar 10 personas para ensayar el fármaco A y se registran los tiempos que tardan en alcanzar un nivel específico en la sangre. El tiempo promedio requerido fue 24.8 min, con una varianza 15.57 min^2 . Al ensayar el fármaco B en 15 personas elegidas al azar, el tiempo promedio fue 22.6 min, con una varianza 17.64 min^2 . La experiencia ha mostrado que los tiempos de absorción de ambos medicamentos se distribuyen normalmente, donde la variabilidad en los tiempos es similar para ambos fármacos. Usando un IC al 95% ¿Es cierta la creencia?

Solución

Suponga que X_1, \dots, X_{10} es una m.a que representa los tiempos de absorción de las 10 personas a las cuales se les administra el fármaco A, $X_i \sim N(\mu_X, \sigma_X^2)$. Análogamente, sea Y_1, \dots, Y_{15} otra m.a que representa los tiempos de absorción de las 15 personas a las cuales se les administra el fármaco B, $Y_j \sim N(\mu_Y, \sigma_Y^2)$.

El enunciado indica que $\sigma_X^2 = \sigma_Y^2$. La información muestral revela que:

$$\bar{x} = 24.8, \quad s_X^2 = 15.57, \quad n = 10,$$

$$\bar{y} = 22.6, \quad s_Y^2 = 17.64, \quad m = 15.$$

Se desea verificar si $\mu_X - \mu_Y > 0$

Así, un IC unilateral superior al 95% para $\mu_X - \mu_Y$, es:

$$(24.8 - 22.6) - t_{0.05,23} S_p \sqrt{\frac{1}{10} + \frac{1}{15}}$$

Usando la información se obtiene que $S_p = \sqrt{16.83} = 4.1024$ y $t_{0.05,23} = 1.714$.

Así el IC es:

$$\left((24.8 - 22.6) - 1.714 \times 4.1024 \sqrt{\frac{1}{10} + \frac{1}{15}}, +\infty \right) \Leftrightarrow (-0.6706, +\infty)$$

Este intervalo permite concluir, que el tiempo promedio de absorción del fármaco B no es menor al tiempo promedio de absorción del fármaco A, con una confianza del 95%.

Ejercicio de IC para $\mu_X - \mu_Y$

El análisis de una m.a. formada por 32 especímenes de acero laminado en frío, dió como resultado una resistencia promedio de 32.5 KSI y una desviación estándar de 4 KSI. Una segunda muestra de 50 especímenes de acero galvanizado mostró una resistencia promedio de 34.2 KSI y una desviación estándar de 5 KSI. ¿Los datos indican que las verdaderas resistencias promedio son diferentes?, para ello construya un IC del 95% para la diferencia de medias.